# Likelihood-based model selection

Population abundance with more than two capture periods

Testing hypotheses about trap responses

# Extending the likelihood-based estimate of population size

- Last time we used ML to estimate population size with two capture periods
  - The probabilities of capture histories were modeled with a single encounter probability
  - Because of this, the probability of histories 01 and 10 had the same predicted frequency
- We are going to extend the method to three capture periods today
  - Eight possible capture histories: 111, 110, 101, 011, 100, 010, 001, 000
- This will allow us to test hypotheses about how our study is working, specifically:
  - Changes in detectability over time
  - Trap response (trap happiness, trap shyness)

# The basic model, $M_0$

| History | t = 1 | t = 2 | t = 3 |
|---|---|---|---|
| 111 | p | p | p |
| 011 | (1-p) | p | p |
| 001 | (1-p) | (1-p) | p |

*This is the model we used for two histories applied to three*

*A single encounter probability, p*

*Focus on these three possibilities – first captures at t = 1, t = 2, or t = 3, and re-captured every time*

# Setup for $M_0$

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Paramete | MLE | Betas | | Multinomial probability of history | | |
| 2 | 001 | 40 | | p | 0.5 | 0 | | 0.125 | | |
| 3 | 010 | 19 | | | | | | 0.125 | | |
| 4 | 011 | 22 | | | | | | 0.125 | | |
| 5 | 100 | 21 | | | | | | 0.125 | | |
| 6 | 101 | 37 | | | | | | 0.125 | | |
| 7 | 110 | 17 | | f(000) | 7.38906 | 2 | | 0.125 | | |
| 8 | 111 | 29 | | | | | | 0.125 | | |
| 9 | 000 | 7.3890561 | | N-hat | 192.389 | | | 0.125 | | |
| 10 | | | | | | | | | | |
| 11 | Mt+1 | Mult coeff. | Prob. Part | LnLikelihood | | | | | | |
| 12 | 185 | 813.71483 | -400.062 | 413.653 | | | | | | |
| 13 | | | | | | | | | | |

*Single encounter probability, p*

*Initial values for betas entered*

*Now to maximize the LnLikelihood by changing the betas*

# Estimates for $M_0$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Paramete | MLE | Betas | | Multinomi | Expected |
| 2 | 001 | 40 | | p | 0.5072 | 0.01439 | | 0.12318 | 25.8235 |
| 3 | 010 | 19 | | | | | | 0.12318 | 25.8235 |
| 4 | 011 | 22 | | | | | | 0.12677 | 26.5778 |
| 5 | 100 | 21 | | | | | | 0.12318 | 25.8235 |
| 6 | 101 | 37 | | | | | | 0.12677 | 26.5778 |
| 7 | 110 | 17 | | f(000) | 24.6489 | 3.20473 | | 0.12677 | 26.5778 |
| 8 | 111 | 29 | | | | | | 0.13048 | 27.3541 |
| 9 | 000 | 24.648877 | | N-hat | 209.649 | | | 0.11968 | 25.0907 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coeff. | Prob. Part | LnLikelihood | | | | | |
| 12 | 185 | 857.7386 | -435.887 | 421.851 | | | | | |

*Encounter probability p = 0.5072*

*Does this fit the data well?*

# Goodness of fit

- We have a probability of each capture history
- We have an estimated population size
- Multiplying the probabilities by the population size gives an expected frequency for each history
- Frequencies observed not well predicted by expected frequencies – not a very good fit

| | |
|---|---|
| N-hat | 209.649 |

| History | Frequency |
|---|---|
| 001 | 40 |
| 010 | 19 |
| 011 | 22 |
| 100 | 21 |
| 101 | 37 |
| 110 | 17 |
| 111 | 29 |
| 000 | 24.648877 |

| Multinomi | Expected |
|---|---|
| 0.12318 | 25.8235 |
| 0.12318 | 25.8235 |
| 0.12677 | 26.5778 |
| 0.12318 | 25.8235 |
| 0.12677 | 26.5778 |
| 0.12677 | 26.5778 |
| 0.13048 | 27.3541 |
| 0.11968 | 25.0907 |

# Mismatch between model and data

| History | Frequency |
|---|---|
| 001 | 40 |
| 010 | 19 |
| 011 | 22 |
| 100 | 21 |
| 101 | 37 |
| 110 | 17 |
| 111 | 29 |
| 000 | 24.648877 |

- With model $M_0$, histories with the same number of 0’s and 1’s have the same probability
- But, we see different frequencies for 001, 010, 100, and for 011, 101, and 110
- Mismatch could be due to:
  - Random chance
  - p that isn’t constant

# Encounter probability can change over time

- Animals can change activity patterns
  - Forage more during the breeding season
  - Call more
- Can have changes in appearance
  - Molting, shedding
- Environmental conditions can change detectability
  - Snow/ice melt
  - Leaves grow or are dropped

*Male*
*American Goldfinches*

*Guess which is non-breeding...*

# Time varying encounter probability, $M_t$

| History | $t = 1$ | $t = 2$ | $t = 3$ |
|---|---|---|---|
| 111 | $p_1$ | $p_2$ | $p_3$ |
| 011 | $(1-p_1)$ | $p_2$ | $p_3$ |
| 001 | $(1-p_1)$ | $(1-p_2)$ | $p_3$ |

*Each capture period has a different encounter probability*

*No response to trapping – subsequent captures are the same as initial*

# Setup for $M_t$

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Parameter | MLE | Betas | | Multinomial probability of history | | | |
| 2 | 001 | 40 | | p1 | 0.26029 | -0.5 | | 0.14242 | | | |
| 3 | 010 | 19 | | p2 | 0.26029 | -0.5 | | 0.14242 | | | |
| 4 | 011 | 22 | | p3 | 0.26029 | -0.5 | | 0.05012 | | | |
| 5 | 100 | 21 | | | | | | 0.14242 | | | |
| 6 | 101 | 37 | | | | | | 0.05012 | | | |
| 7 | 110 | 17 | | f(000) | 7.38906 | 2 | | 0.05012 | | | |
| 8 | 111 | 29 | | | | | | 0.01763 | | | |
| 9 | 000 | | | N-hat | 192.389 | | | 0.40475 | | | |
| 10 | | | | | | | | | | | |
| 11 | Mt+1 | Mult coeff | Prob. Port | LnLikelihood | | | | | | | |
| 12 | 185 | 813.715 | -507.2 | 306.515 | | | | | | | |
| 13 | | | | | | | | | | | |

*Different encounter probability for each capture period*

*I made up the data to be time-varying, so we know this is the right model*

# Estimates for $M_t$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Parameter | MLE | Betas | | Multinom | Expected |
| 2 | 001 | 40 | | p1 | 0.50082 | 0.00165 | | 0.17878 | 37.1255 |
| 3 | 010 | 19 | | p2 | 0.41896 | -0.1628 | | 0.08022 | 16.6592 |
| 4 | 011 | 22 | | p3 | 0.6164 | 0.23495 | | 0.12891 | 26.7691 |
| 5 | 100 | 21 | | | | | | 0.11163 | 23.1805 |
| 6 | 101 | 37 | | | | | | 0.17937 | 37.2479 |
| 7 | 110 | 17 | | f(000) | 22.6581 | 3.12052 | | 0.08049 | 16.7142 |
| 8 | 111 | 29 | | | | | | 0.12933 | 26.8574 |
| 9 | 000 | | | N-hat | 207.658 | | | 0.11126 | 23.1043 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coeff | Prob. Port | LnLikelihood | | | | | |
| 12 | 185 | 853.441 | -423.394 | 430.047 | | | | | |

*Expected frequencies look better*

*Now histories with same number of 1's have different expected freqs*

# Trap response

- Behavioral response to trapping can be:
  - Trappy happy = increased chance of detection after first capture (usually attraction to baited traps)
  - Trap shy = decreased chance of detection after first capture (aversion to traps due to bad experience)
- Use of a different survey method after initial capture can also affect encounter probability
  - First capture is an actual capture so that a mark can be applied
  - Subsequent captures are re-sightings from a distance
- We need to know if this is happening, and if so account for it

# Trap response model, $M_b$

| History | t = 1 | t = 2 | t = 3 |
|---|---|---|---|
| 111 | p | c | c |
| 011 | (1-p) | p | c |
| 001 | (1-p) | (1-p) | p |

*Initial captures are the same probability at each capture period (p)*

*But, after the first capture the probability changes (c)*

*What would the probability of history 010?*

# Setup for $M_b$

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Paramete | MLE | Betas | | Multinomial probability of history | | |
| 2 | 001 | 40 | | p | 0.26029 | -0.5 | | 0.14242266 | | |
| 3 | 010 | 19 | | c | 0.26029 | -0.5 | | 0.14242266 | | |
| 4 | 011 | 22 | | | | | | 0.05011513 | | |
| 5 | 100 | 21 | | | | | | 0.14242266 | | |
| 6 | 101 | 37 | | | | | | 0.05011513 | | |
| 7 | 110 | 17 | | f(000) | 7.38906 | 2 | | 0.05011513 | | |
| 8 | 111 | 29 | | | | | | 0.01763431 | | |
| 9 | 000 | | | N-hat | 192.389 | | | 0.40475232 | | |
| 10 | | | | | | | | | | |
| 11 | Mt+1 | Mult coef | Prob. Port | LnLikelihood | | | | | | |
| 12 | 185 | 813.715 | -507.2 | 306.515 | | | | | | |

*Up to first capture, encounter probability is p, probability of no capture is (1-p)*

*After first capture, encounter probability is c, probability of no capture is (1-c)*

# Estimates for $M_b$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Paramete | MLE | Betas | | Multinomial p | Expected |
| 2 | 001 | 40 | | p | 0.42781 | -0.14489 | | 0.14006595 | 31.8154 |
| 3 | 010 | 19 | | c | 0.53815 | 0.07638 | | 0.11305491 | 25.6799 |
| 4 | 011 | 22 | | | | | | 0.13173338 | 29.9227 |
| 5 | 100 | 21 | | | | | | 0.09125282 | 20.7277 |
| 6 | 101 | 37 | | | | | | 0.10632924 | 24.1522 |
| 7 | 110 | 17 | | f(000) | 42.1456 | 3.74113 | | 0.10632924 | 24.1522 |
| 8 | 111 | 29 | | | | | | 0.12389652 | 28.1426 |
| 9 | 000 | | | N-hat | 227.146 | | | 0.18733796 | 42.553 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coef | Prob. Port | LnLikelihood | | | | | |
| 12 | 185 | 890.568 | -467.087 | 423.48 | | | | | |
| 13 | | | | | | | | | |

*Better than $M_0$, but a couple of misfires – 110 and 101 have same expected frequencies, but very different observed freqs*

# The real model

- These are contrived data – random data generated with:
  - Probability of capture of 0.5 for the first interval
  - Probability of capture of 0.4 for the second interval
  - Probability of capture of 0.6 for the third
- The model that came closest to these values was $M_t$, and it had the highest log likelihood
- Would we reach the correct conclusion with our analysis, and pick this model? How do we compare models?

# Comparing models

- $M_0$, $M_t$, $M_b$ represent three different hypotheses about encounter probability
- All are “wrong”, in that all are incomplete
  - In a contrived example, mismatches are due just to random variation
  - In real data mismatches can also be due to unmeasured factors, only treated as random variation because we don’t have information about them
- But, we should prefer to interpret the model that is most consistent with the data – we can use the **Method of Support** to tell us
  - Support for a model is measured with its likelihood
  - The model best supported by the data will be the one we interpret

# Likelihood as a measure of **support**

- Each capture model is at best an approximation of what is really happening = the True Model
  - The TM is a complete explanation for every 0 and 1
  - So many factors determine every 0 and 1 that the TM is essentially unknowable
  - But, our models approximate TM, and we should prefer models that are closest to the TM
- We will use "Akaike's Information Criterion", to decide which of our approximations is best supported by the data
  - AIC is a measure of relative distance to the TM
  - We don't know the TM for AIC to tell us which model is closest to TM

$$AIC = -2\ln\left(L\left(model \mid the\ data\right)\right) + 2K$$

*K = number of estimated parameters*

# Properties of AIC

- Only interpretable in comparison with other AIC values
  - A relative measure of distance from TM, so smaller is better
- Every model compared has to use the same response data (i.e. the same set of frequencies of capture histories)
- Balances model fit and model complexity
  - -2ln(L(model|data)) gets smaller the better the model fits the data
  - 2K gets bigger the more parameters are added (i.e. more complex the model is)
- Smallest AIC is obtained with a simple model that fits the data well

$$AIC = \overbrace{-2\ln\left(L\left(model \middle| the\, data\right)\right)}^{\textit{Model fit}} + \overbrace{2K}^{\textit{Model complexity}}$$

# AIC values for our three models

| Model | -2lnLikelihood | 2K | AIC |
|---|---|---|---|
| M0 | -843.7023538 | 4 | -839.702 |
| Mt | -860.0937198 | 8 | -852.094 |
| Mb | -846.9608798 | 6 | -840.961 |

*Which should be smallest?*

*Which is smallest?*

# Refinements to AIC

- For sample sizes with less than 40 observations per parameter ($n/K < 40$), use $AIC_c$

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$$

*n is the total number of captures*

- $AIC_c$ is AIC with an additional penalty applied for complex models – the size of penalty is greater when n is small
  - Complex models become less reliable when sample sizes are small
  - Thus, adding more parameters should cost more when the sample size is small

# $AIC_c$ for our models

| Model | -2lnLikelihood | 2K | AIC | AICc |
|---|---|---|---|---|
| M0 | -843.7023538 | 4 | -839.702 | -839.602 |
| Mt | -860.0937198 | 8 | -852.094 | -851.755 |
| Mb | -846.9608798 | 6 | -840.961 | -840.759 |

*Like likelihoods, AIC values are only meaningful in comparison to other AIC values*

*We can make the comparison easier...*

# $\Delta$AIC

- To make comparison among $AIC_c$ values easier, subtract the smallest $AIC_c$ from all candidates under consideration = $\Delta AIC_c$'s
  - Best supported will have $\Delta AIC_c$ of 0
  - $\Delta AIC_c$ between 4 and 7 indicate substantially reduced support for a model relative to the best-supported
  - $\Delta AIC_c$ greater than 10 indicates the data doesn’t support the model at all

# $\Delta AIC_c$ for our models

| Model | -2lnLikelihood | 2K | AIC | AICc | ΔAIC |
|---|---|---|---|---|---|
| M0 | -843.7023538 | 4 | -839.702 | -839.602 | 12.15238 |
| Mt | -860.0937198 | 8 | -852.094 | -851.755 | 0 |
| Mb | -846.9608798 | 6 | -840.961 | -840.759 | 10.99554 |

*$M_t$ is the best supported, and the other two have essentially no support*

*Based on these $\Delta AIC_c$ values, we really only have to consider $M_t$ as a reasonable explanation for the data*

*$M_t$ is the model that generated the data, so $AIC_c$ worked!*

# Akaike weights

- Allow us to quantify model uncertainty
- If we collected new data, what are the chances that each model would be best supported?
- Based on the DAIC($_c$)'s of a set of models under consideration
- D's that are in the "unsupported" range will have very low weights

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum \exp\left(-\frac{1}{2}\Delta\right)}$$

# Weights for our models

| Model | -2lnLikelihood | 2K | AIC | AICc | ΔAIC | w |
|---|---|---|---|---|---|---|
| M0 | -843.7023538 | 4 | -839.702 | -839.602 | 12.15238 | 0.002 |
| Mt | -860.0937198 | 8 | -852.094 | -851.755 | 0 | 0.994 |
| Mb | -846.9608798 | 6 | -840.961 | -840.759 | 10.99554 | 0.004 |

*Mt is the best supported, and we expect to select it 99.4% of the time if we repeated the experiment*

# Interpreting $M_t$

- Encounter probability lowest at the second trapping period, highest in the third
- The population estimate is 207.658
- There were only 22.65 individuals that were not observed

| D | E |
|---|---|
| Parameter | MLE |
| p1 | 0.50082 |
| p2 | 0.41896 |
| p3 | 0.6164 |
| | |
| | |
| f(000) | 22.6581 |
| | |
| N-hat | 207.658 |

# Data with a trap response

- New contrived data set
  - Probability of first capture is 0.6
  - After first capture, probability of capture changes to 0.15 (happy, or shy?)
- We will fit the three models and see how they compare

## $M_0$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Parameter | MLE | Betas | | Multinom | Expected |
| 2 | 001 | 17 | | p | 0.18165 | -0.69021 | | 0.12165 | 50.0039 |
| 3 | 010 | 41 | | | | | | 0.12165 | 50.0039 |
| 4 | 011 | 5 | | | | | | 0.027 | 11.0995 |
| 5 | 100 | 93 | | | | | | 0.12165 | 50.0039 |
| 6 | 101 | 14 | | | | | | 0.027 | 11.0995 |
| 7 | 110 | 13 | | f(000) | 225.044 | 5.41629 | | 0.027 | 11.0995 |
| 8 | 111 | 3 | | | | | | 0.00599 | 2.46379 |
| 9 | 000 | | | N-hat | 411.044 | | | 0.54804 | 225.27 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coeff. | Prob. Part | LnLikelihood | | | | | |
| 12 | 186 | 1069.3466 | -584.367 | 484.98 | | | | | |

## $M_t$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Parameter | MLE | Betas | | Multinom | Expected |
| 2 | 001 | 17 | | p1 | 0.33537 | -0.33553 | | 0.05873 | 21.539 |
| 3 | 010 | 41 | | p2 | 0.16905 | -0.72336 | | 0.10041 | 36.8256 |
| 4 | 011 | 5 | | p3 | 0.10634 | -0.90646 | | 0.01195 | 4.38182 |
| 5 | 100 | 93 | | | | | | 0.24904 | 91.339 |
| 6 | 101 | 14 | | | | | | 0.02963 | 10.8683 |
| 7 | 110 | 13 | | f(000) | 180.764 | 5.19719 | | 0.05066 | 18.5817 |
| 8 | 111 | 3 | | | | | | 0.00603 | 2.211 |
| 9 | 000 | | | N-hat | 366.764 | | | 0.49355 | 181.018 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coeff | Prob. Port | LnLikelihood | | | | | |
| 12 | 186 | 1040.53 | -524.864 | 515.662 | | | | | |

## $M_b$

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | History | Frequency | | Parameter | MLE | Betas | | Multinomial p | Expected |
| 2 | 001 | 17 | | p | 0.63456 | 0.27248 | | 0.08474247 | 16.528 |
| 3 | 010 | 41 | | c | 0.13014 | -0.83266 | | 0.20171513 | 39.3422 |
| 4 | 011 | 5 | | | | | | 0.03017788 | 5.88584 |
| 5 | 100 | 93 | | | | | | 0.48014878 | 93.6474 |
| 6 | 101 | 14 | | | | | | 0.07183335 | 14.0102 |
| 7 | 110 | 13 | | f(000) | 9.03822 | 2.20146 | | 0.07183335 | 14.0102 |
| 8 | 111 | 3 | | | | | | 0.01074673 | 2.09602 |
| 9 | 000 | | | N-hat | 195.038 | | | 0.04880231 | 9.51831 |
| 10 | | | | | | | | | |
| 11 | Mt+1 | Mult coeff | Prob. Port | LnLikelihood | | | | | |
| 12 | 186 | 824.104 | -305.325 | 518.779 | | | | | |

# Which is best supported?

| Model | -2lnLikelihood | 2K | AIC | AICc | ΔAIC | w |
|---|---|---|---|---|---|---|
| M0 | -969.9601712 | 4 | -965.96 | -965.86 | 65.49624 | 0.000 |
| Mt | -1031.324337 | 8 | -1023.32 | -1022.99 | 8.371057 | 0.015 |
| Mb | -1037.558092 | 6 | -1031.56 | -1031.36 | 0 | 0.985 |

# Interpreting $M_b$

- Encounter probability decreased after the first capture (trap shy)
- There were 9 animals that were never seen
- The population estimate was 195

| D | E | F |
|---|---|---|
| Parameter | MLE | Betas |
| p | 0.63456 | 0.27248 |
| c | 0.13014 | -0.83266 |
| | | |
| | | |
| | | |
| f(000) | 9.03822 | 2.20146 |
| | | |
| N-hat | 195.038 | |

# What if multiple models are well supported?

- The results are not always this clear
  - $M_b$ and $M_t$ don't always yield distinctly different results, because there is a time component to $M_b$
  - With smaller sample sizes, the more complex $M_t$ and $M_b$ models may not have sufficiently big differences in likelihood to be better supported than $M_0$
- If more than one model is well supported, say so!
  - Separating the well-supported from the poorly supported models is worthwhile, even if two or more are well supported
  - If the data don't differentiate between the competing hypotheses, you don't want to pretend otherwise

# FYI: time varying with trap response ($M_{tb}$)

| History | t = 1 | t = 2 | t = 3 |
|---|---|---|---|
| 111 | $p_1$ | $c_2$ | $c_3$ |
| 011 | $(1-p_1)$ | $p_2$ | $c_3$ |
| 001 | $(1-p_1)$ | $(1-p_2)$ | $p_3$ |

*It is also possible to have a combination of time-varying encounter probability, and a trap response*

*This is the most complex model, with 5 parameters, and it's difficult to fit with only three capture periods...we won't work with this one*